# 24.962 Phonology Squib: English Shm-Reduplication in Optimality Theory

Ezra Keshet

May 10, 2004

## 1 Introduction

Nevins & Vaux (2003) present the results of a survey on shm-reduplication and argue against an Optimality Theory account of this data. The most interesting result of the survey is the degree of variation in answers. Below, I present an OT analysis of the same data, including simple re-rankings of constraints to explain variations.

## 2 Data

Canonical cases of shm-reduplication include constructions such as "table-shmable" and "bagel-shmagel." As revealed in the Nevins/Vaux survey, most people agree with the shm-reduplicated forms of these words. Differences arise, however, when any of these conditions hold (examples from Nevins & Vaux):

- Liquids in the onset: broom, breakfast, street, rich, floss, floozie
- Glides in the onset: union, use, confusion, wig, dwarf
- Stridents elsewhere in the word: witches, Ashmont, ash, massage, schnozz
- Non-initial stress: confusion, obscene, massage, terrific, arcade

Although in most cases above the majority of respondents could agree on a reduplicated form, some words revealed significant minority judgments. For instance, nearly as many people chose "broom-shmroom" as chose "broom-shmoom."

## 3 Analysis

Below I will present Optimality Theory constraints and rankings to explain shm-reduplication as outlined above. For each case, I will present several rankings of the same constraints, each ranking representing the different attested outputs,

as reported in the survey cited above. For the purposes of this section, I am only concerned with capturing all the possible variations of this phenomenon – not with their varying degrees of prevalence. The main intuition is that variations occur when a particular phonotactic (markedness) constraint is re-ranked with respect to a particular faithfulness constraint.

## 3.1 Canonical Case

Tableau 1 shows the normal case of shm-reduplication. If such simple cases as these are the only ones examined, two well-known constraints will suffice:

(1) table-shmable, Alderete et al., Tableau 12:

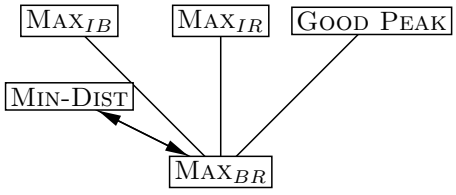
| | /tejbəl-ʃm+RED/ | MAX$_{IO}$ | MAX$_{BR}$ |
|---|---|---|---|
| ☞ | tejbəl-ʃmejbəl | | t |
| | tejbəl-tejbəl | ʃm! | |
| | ʃmejbəl-tejbəl | t! | ʃm |
| | ʃmejbəl-ʃmejbəl | t! | |

I remain ambivalent about the different models of reduplication. For clarity, though, I will adopt the following model: I represent the affix as /-ʃm+RED/. In the output forms, the letters of the base are linked to the letters of the input word (base) and the letters of the reduplicant are either also linked to the input base. So, in the tableau above, tejbəl-ʃmejbəl wins because it optimally preserves the base (satisfying MAX$_{IO}$). As we will see below, however, this is much too simple an analysis to work for all cases.

## 3.2 Liquids

As mentioned above, speakers differ when forming the shm-reduplicated form of words where the relevant onset contains a liquid. For instance, 37% of respondents allowed the form "broom-shmroom."

(2) Hasse diagram (↔ indicates constraints that are ranked differently for different speakers):



(3) broom-shmoom:

| | /brum-ʃm+RED/ | MAX$_{IB}$ : GOOD PEAK : MIN-DIST | MAX$_{BR}$ |
|---|---|---|---|
| ☞ | brum-ʃmum | : : | br |
| | brum-ʃmrum | : : *! | b |
| | brum-ʃmbrum | : *! : | |
| | um-ʃmum | br! : : | |

(4) broom-shmroom:

| | /brum-ʃm+RED/ | MAX$_{IB}$ : GOOD PEAK | MAX$_{BR}$ | MIN-DIST |
|---|---|---|---|---|
| | brum-ʃmum | : | br! | ʃm |
| ☞ | brum-ʃmrum | : | b | * |

Tableaux 3 and 4 show two variations of reduplication on the word "broom". These variations can be explained by a simple re-ranking of two constraints: MIN-DIST SONORITY and MAX$_{BR}$. MIN-DIST here assigns violations for each pair of adjacent consonants that are within two categories of each other in the sonority hierarchy. The hierarchy I presume is shown in 5:

(5) stops ⟩⟩ fricatives ⟩⟩ nasals ⟩⟩ (empty) ⟩⟩ liquids ⟩⟩ glides

Under this constraint, English onsets will disallow combinations such as "bf," "bm," "fm," and "mr"[1] but allows ones like "br" and "fy". Since nasals can occur next to glides and frications next to liquids, I have added an empty position in the sonority hierarchy. It seems that in some relevant dimension, the nasals are farther from the liquids than, for instance, the stops are from the nasals. This extra distance allows combinations such as "fl" and "my" to occur. The general constraint as formulated above is merely an approximation of this actual (yet-to-be-discovered) distance metric. One notable exception to this generalization is "s": it can violate this constraint: "sm," "sn," and even "st" are all allowed. Hopefully, the analysis for these cases will extend to ʃto allow ʃm in reduplicated words. The other markedness constraint, GOOD PEAK, assigns violations to each sonority peak in a syllable other than the nucleus.

I have also proposed more specific faithfulness constraints. MAX$_{IB}$ assigns violations to structures where the base of a reduplicated form is missing elements. Since this never happens in this process, MAX$_{IB}$ is a highly ranked constraint. MAX$_{IR}$ assigns violations to structures where the output form of a reduplicant is missing elements present in its input form. This constraint is also unviolated in shm-reduplication. I have proposed these specific constraints to replace the more general MAX$_{IO}$ due to the following circumstances. In tableau 4 above, MAX$_{BR}$ must outrank MIN-DIST; and as we saw in tableau 1, MAX$_{IO}$ must always outrank MAX$_{BR}$. However, if the general MAX$_{IO}$ ever outranked MIN-DIST, as it must in this situation by transitivity, we would expect to find sequences such as "mr" in the general vocabulary, and not just in specialized

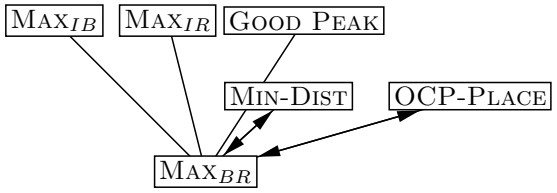
[1] If "mr" is disallowed in an onset, you might expect "rm" to be disallowed in a coda. I account for the occurrence of "rm" as in "harm" by claiming that the "r" is part of a diphthong and not strictly a consonant. Evidence for this includes the fact that "rm" always follows a vowel, and only certain vowels, at that.

situations such as shm-reduplicaton. I have not included MAX$_{IR}$ because it is not relevant for the candidates shown.

### 3.3 Glides

19% of Nevins and Vaux's respondents chose the form [junjən-ʃmjunjən] for the reduplicated form of "union." 5% chose [wɪg-ʃmwɪg] for "wig." The analysis of these forms with glides could be captured with a *CCC constraint (perhaps just for onsets, perhaps for whole words). However, such a constraint would also rule out forms that preserve a liquid onset, such as "broom-shmroom," and a small percentage of respondents who allowed "broom-shmroom" did not allow "union-shmyoonion." Unless this data is just noise in the system (see below), this fact points to an analysis where separate constraints rule out the preservation of liquids and the preservation of glides. For glides, I tentatively propose that the same constraint that rules out "ty," "ny," "sy" and "ʃy" rules out "ʃmy." OCP-PLACE-ONSET disallows two consonants with the same place of articulation from coocurring in an onset. So, "my" is allowed by this constraint, as in "music," but "ʃy," "ʃmy," and "mw" are ruled out. This is true at least for American English in words one might expect to find these sequences: "tissue" and "eschew."[2]

(6) Hasse:




(7) union-shmoonion:

| | /junjən-ʃm+RED/ | OCP-PLACE | MAX$_{BR}$ |
|---|---|---|---|
| ☞ | junjən-ʃmunjən | | j |
| | junjən-ʃmjunjən | *! | |

(8) union-shmyoonion:

| | /junjən-ʃm+RED/ | MAX$_{BR}$ | OCP-PLACE |
|---|---|---|---|
| | junjən-ʃmunjən | j! | ʃm |
| ☞ | junjən-ʃmjunjən | | * |

For glides, Nevins and Vaux argue that the ambiguity between speakers is not a re-ranking of the constraints, but rather whether the speaker analyzes the glide as part of the nucleus. In the framework of this paper, glides in the nucleus would not be subject to the relevant markedness violations that glides in

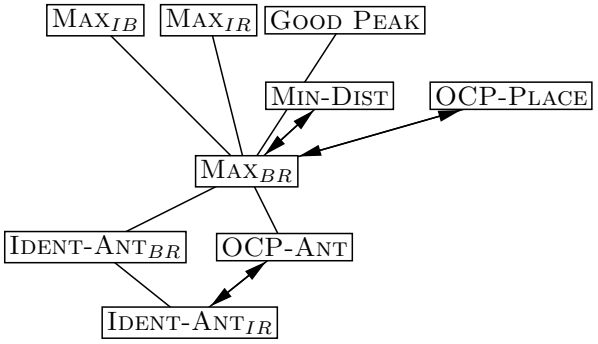
[2]It would be interesting to hear how British English speakers would form the shm-reduplicated form of "union."

the onset are subject to. This analysis has trouble explaining the speakers who preserved liquids but not glides: even speakers who believe the glide is outside the nucleus should preserve it if they are preserving the r in "breakfast."

## 3.4 Stridents

Some speakers change the "ʃ" in the affix to an "s" when other stridents are present in the word. For instance, 17% of respondents allowed "Ashmont-smashmont." In this case, it seems that an OCP-ANTERIOR constraint outranks an IDENT-ANTERIOR$_{IR}$ constraint. In the tableau below I illustrate this (minority) ranking:

(9) Hasse:




(10) Ashmont-smashmont:

| /aʃmɔnt-ʃm+RED/ | IDENT-ANT$_{BR}$ : OCP-ANT | IDENT-ANT$_{IR}$ |
|---|---|---|
| ☞ aʃmɔnt-smaʃmɔnt | : | * |
| aʃmɔnt-ʃmaʃmɔnt | : *! | |
| aʃmɔnt-ʃmasmɔnt | *! : | * |

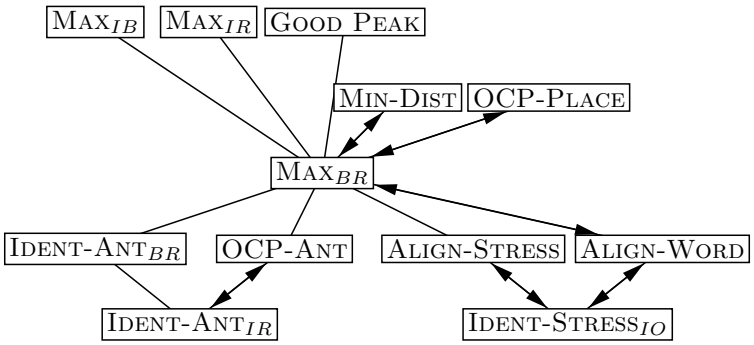
Notice that IDENT-ANT$_{BR}$ ensures that only the ʃin the input to the reduplicant changes, not those copied from the base.

## 3.5 Non-initial Stress

The variations found in shm-reduplicated words with non-initial stress are slightly more complex. Some speakers allow the "shm" affix to attach to the unstressed first syllable. Others have two basic repair strategies: move the affix (on the reduplicant) or move the stress (on the base and reduplicant). Once again, these variations are explainable in terms of a re-ranking of constraints. ALIGN-WORD = ALIGN(ʃm, L, PrWd, L) assigns violations for every segment between the left edge of the affix and the left edge of the reduplicant. ALIGN-STRESS = ALIGN(ʃm, L, $\acute{\sigma}$, L) assigns violations for every segment between the left edge of

the affix and the left edge of a stressed syllable. IDENT-STRESS$_{IO}$ assigns violations for elements (such as stress) that have changed or moved when comparing the input and output.

(11) Hasse:




(12) óbscene-shmóbscene:

| /ʊbsín-ʃm+RED/ | ALIGN-STRESS : ALIGN-WORD | IDENT-STRESS$_{IO}$ |
|---|---|---|
| ☞ ʊ́bsin-ʃmʊ́bsin | : | * |
| ʊbsín-ʃmʊbsin | *! : | |
| ʊbsín-ʊbʃmín | : *! | |

(13) obscéne-shmobscéne:

| /ʊbsín-ʃm+RED/ | ALIGN-WORD : IDENT-STRESS$_{IO}$ | ALIGN-STRESS |
|---|---|---|
| ʊ́bsin-ʃmʊ́bsin | : *! | |
| ☞ ʊbsín-ʃmʊbsín | : | * |
| ʊbsín-ʊbʃmín | *! : | |

(14) obscéne-obshméne:

| /ʊbsín-ʃm+RED/ | ALIGN-STRESS : IDENT-STRESS$_{IO}$ | ALIGN-WORD |
|---|---|---|
| ʊ́bsin-ʃmʊ́bsin | : *! | |
| ʊbsín-ʃmʊbsin | *! : | |
| ☞ ʊbsín-ʊbʃmín | : | * |

Some respondents preferred to drop the first syllable of a word like this entirely, forming "obscene-shmene". This form requires reranking ALIGN-WORD above MAX$_{BR}$ (which is assumed to be just above ALIGN-WORD in tableau 14). "shmene" satisfies ALIGN-WORD, but has more violations for MAX$_{BR}$.

Nevins and Vaux note the possible stress shift in the base in section 5.4, but mention that it is "not clear whether this is due to the rhythm rule or a desire for prosodic identity between the base and the reduplicant." Longer forms showing this stress shift, such as "cónfusion-shmónfusion" rule out the rhythm rule possibility, leaving, as analysed above, a desire for prosodic identity between

the base and reduplicant. This is surprising, hence interesting, given the relative scarcity of examples exhibiting Input-to-Base Correspondance constraints being outranked. Such an effect may argue against theories that predict no backcopying in reduplication, such as Morphological Doubling Theory.

# 4 Discussion

In this section, I will raise some more thought-provoking questions and answer (some of) them.

## 4.1 Nevins & Vaux

Nevins and Vaux raise some potential problems with the Optimality Theory account of shm-reduplication. In this section, I will address these problems and raise some problems for their analysis of the phenomenon.

In section 5, Nevins and Vaux argue against Alderete et al. (1999), claiming that the latter do not explain the full set of variations exhibited by shm-reduplication, for instance "wig-shmig" versus "wig-shmwig". Hopefully the constraints, rankings, and re-rankings above go a long way to explaining this variation. Nevins and Vaux also raise some specific problems, though. For instance, they claim that under the ranking $\textsc{Max}_{IO} \rangle\rangle \textsc{Max}_{BR}$ proposed above (as in 1), the optimal output for "eel" would be "shmeel-shmeel". However, under the assumptions mentioned above, the tableau for this input is actually as below:

(15) "eel-shmeel":

| | /i$_1$l$_2$-ʃ$_3$m$_4$+RED/ | $\textsc{Dep}_{IB}$ | $\textsc{Dep}_{BR}$ | Uniformity |
|---|---|---|---|---|
| ☞ | i$_1$l$_2$-ʃ$_3$m$_4$i$_1$l$_2$ | | ʃm | il |
| | ʃ$_3$m$_4$i$_1$l$_2$-ʃ$_3$m$_4$i$_1$l$_2$ | ʃm! | | ʃmil |

$\textsc{Dep}_{IB}$ penalizes segments in the output base not in the input base. $\textsc{Dep}_{BR}$ penalizes segments in the reduplicant not linked to those base. Uniformity penalizes single segments in the input linked to multiple segments in the output. Under this ranking, the correct output form is chosen.

I agree with Nevins & Vaux that the variation in shm-reduplication is not explainable by a re-ranking of the Max constraints – but that is not how I do it above. Instead, it is the interation of the Max constraints and phonotactic constraints that causes the variation.

I do have one or two points to raise about the system proposed by Nevins & Vaux, based on "Anchor Points". They claim that reduplication can only target one of the following landmarks:

(16) Anchor Points: 1st syllable, 1st foot, 1st consonant, 1st vowel, stressed syllable, final syllable. ($\approx 6$)

Crucially, reduplication cannot count any items – so it cannot target the second syllable, for instance. Nevins & Vaux do relax the rules somewhat, later

in the paper, to allow rules to refer to the first onset, the first nucleus, and the first nuclear segment. However, forms such as "street-shmreet" pose a problem for this system. There is no good way to write a rule without counting. Either reduplication replaces the first two consonants or it replaces everything up to the first liquid. The first formulation involves counting and the second does not work for words without liquids.

## 4.2 Variations in the Variations

At least two interesting questions remain: why is there variantion among speakers and why is there variation among words (e.g., different example words in Nevins & Vaux)?

On the first point, I basically agree with Nevins and Vaux (translated into my framework). Speakers first hear shm-reduplication with only canonical cases (e.g. "table-shmable"). These forms are ambiguous between the rankings shown above. Therefore, different speakers form different opinions as to the rankings, resulting in the variations shown.

This sort of system may be learnable using the Gradual Learning Algorithm (Boersma & Hayes 1999). For instance, after hearing many English words, a speaker has ranked and ranking the MIN-DIST constraint many times. Whether it has come to rest above or below $\text{MAX}_{BR}$ in the overall hierarchy is what determines if (or how often) the speaker will preserve liquids in shm-reduplication. The details of how to apply the GLA in cases where very little data exists (like for $\text{IDENT-ANT}_{IR}$) is still an open question. Perhaps the fact that most people agree on a form shows an inherent preference for certain rankings of constraints (i.e., an initial ranking). Also, one could argue that cases that lower or raise a related constraint, such as $\text{IDENT-ANT}_{IO}$ might also perturb the ranking of $\text{IDENT-ANT}_{IR}$ by analogy. To really explore these issues in more depth would require a much more careful study, including multiple forms exhibiting the same variations to test for free variation and careful correlation of the answers of single respondents.

As for the second question, I have a few things to say – mostly based on further phonotactic constraints. For instance, although "broom-shroom" is almost as prevalent as "broom-shmoom", "rich-shmich" is much preferred over "rich-shmrich". Same for "dwarf-shmwarf" (31%) and "dwarf-shmarf" (61%) versus "wig-shmwig" (5%) and "wig-shmig" (92%). These are due to a general difference between liquids and glides in clusters and on their own. The "mw" cluster is especially dispreferred because of the shared place of articulation (lips). There also seems to be a dispreference for extremely short words in the construction. These cases may also show the interaction of constraints in a stochastic manner.

One of the most interesting points of the Nevins and Vaux paper is the explanation of the limited range of possible variations in shm-reduplication , based on possible "Anchor Points." However, as shown above, this characterization is insufficient to cover all the facts. But, does the system sketched above explain the limited range of variations? First, let's assume that the only variation in

speakers occurs where marked with a double-arrow in the Hasse diagrams above; other constraint rankings are well defined (using the GLA) over the corpus of English data available to a learner. In this case, all of the possible variations predicted do occur, as shown in the reranked tableaux above. So, as long as you assume the distribution of constraints as proposed above, the range of variation is still limited.

# 5 Conclusion

In this paper, I have shown that a small set of phonotactic constraints, in combination with a small set of faithfulness constraints can explain the variation in shm-reduplication as described in Nevins & Vaux (2003). In fact, this system can explain a case not satisfactorily covered by the Anchor Points theory: "street-shmreet". Questions still remain, such as preferences for certain variations and how such low-ranked constraints might be learned. However, the intuition that phonotactic constraints drive the formation of shm-reduplicants seems to hold true.

# 6 References

Alderete, John, Benau, Jill, Benua, Laura, Gnanadesikan, Amalia, McCarthy, John, and Urbanczyk, Suzanne. 1999. Reduplication and Segmental Unmarkedness. Manuscript, University of Massachusetts, Amherst. [ROA 134-0996]

Boersma, Paul and Hayes, Bruce. 1999. Empirical Tests of the Gradual Learning Algorithm. [ROA 348-1099]

Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (Kingston, John and Beckman, Mary, eds.). Cambridge University Press.

Inkelas, Sharon and Zoll, Cheryl. Reduplication as Morphological Doubling. Cambridge University Press (to appear).

Nevins, Andrew and Vaux, Bert. 2003. Metalinguistic, Shmetalinguistic: The Phonology of Shm-Reduplication. Presented at the Chicago Linguistics Society, April 2003. [http://web.mit.edu/anevins/www/schm.pdf]